

# Statistical Computing in the United States

by Roy C. Milton\*

**Recent history and developments related to the increase in statistical computing activities in the United States and by U.S. participants in international efforts are reviewed, with emphasis on important events, organizations, references, and products which contribute to informed selection and use of statistical programs. Three features matrices for major statistical packages are included as potential aids to Japanese statisticians in assessing the utility of these packages in biostatistical applications.**

## Introduction

Computers and statistical computing play an increasingly important role in statistical analysis of data in many areas of application, and biostatistics is no exception. The computer is a primary tool for most modern statisticians, for reasons of convenience, necessity, or both. It is convenient as it shortens the time and effort required for computations, enabling almost concurrent progress of numeric and analytic processes in the statistical evaluation of data, thus enhancing the interaction between data and ideas. It is necessary, as it permits computations of a scope or complexity that was often considered but not routinely feasible only 10 to 20 years ago. The availability of computer programs or packages for statistical analyses establishes to some extent de facto standards for the choice and application of statistical methodology to the analysis of data.

In recent years the statistical profession has become increasingly aware of the importance of these developments and of the need to monitor, guide, and evaluate them. Appropriate and adequate mechanisms to meet this need are evolving even as statistical computing itself continues to develop, and there are signs that the gap between ongoing development and critical evaluation is at least growing more slowly, if not actually narrowing. Major efforts are now underway, both in individual countries and internationally, toward reducing this gap.

This presentation is (1) a summary of recent history and developments in statistical computing

primarily in the U.S., with special emphasis on availability and evaluation of statistical program packages, and (2) a guide to current, important references and resources which may assist a prospective user of statistical programs in becoming aware of activities and reports relevant to a more informed selection and use. Although references and resources to be mentioned are of quite general interest in terms of areas of statistical application, those included are usually relevant to the broad interests of biostatisticians if not also to their sometimes specialized interests. Some references are excluded from this presentation either somewhat arbitrarily or as perhaps too distant from biostatistical applications.

## A View of Recent History

This brief review of the history and development of statistical computing is selective and meant to be informative rather than exhaustive.

By the early 1960's, the development and availability of computer hardware and software had encouraged individuals and groups to produce general statistical programs and packages for routine use in data analysis. Some of the pioneers in this effort are still prominent contributors to statistical computing, through continuing dedication, perseverance and skill. Professional recognition and motivation for these activities, necessary for their further development, has often been erratic and inadequate, with improvement generally slow to occur but increasing in the present decade.

Three major events in the late 1960's marked the beginnings of today's organized professional interest

---

\*Office of Biometry and Epidemiology, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20205, U.S.

in statistical computing by statisticians and their societies. The first Computer Science and Statistics: Symposium on the Interface was held in 1967, beginning a series of annual symposia which have gained wide participation by statisticians and sponsorship by statistical societies. The purpose of the symposia has been described as: to facilitate the design and improvement of computing systems and software by use of statistics; to facilitate the implementation of statistical methods by making the most efficient use of computation machinery; to enhance the combined use of statistics and computation as adjuncts to the basic and applied sciences. Mann (1) provides a detailed history of the symposia.

Later in 1967, at the meeting of the International Statistical Institute (ISI) in Sydney, presentations and discussions took place which led to the organization of the Conference on Statistical Computation, held at the University of Wisconsin in 1969, to "... present and evaluate the current status of some basic aspects of the organization of statistical data processing and computing, and suggest directions for future research and development" (2). The problems posed at that meeting — duplication of effort, communication between statistical programs, definition and specification of data structures, and statistical processing languages — continue among those being studied today.

The third event was the recognition by the American Statistical Association of the role of computers in statistics, by organizing sessions on the topic beginning with the 1967 annual meeting, followed by creation of an ad hoc Committee on Computers in Statistics.

These rather modest events were important because they provided the basis for more substantive developments in the 1970's. The Symposia on the Interface and the ASA sessions grew to provide an effective forum for presentation and recognition of work by professional statisticians in statistical computing. The ad hoc ASA committee was succeeded in 1972 by the Section on Statistical Computing, which has become increasingly active in pursuing the areas outlined in its charter (3) as follows:

#### **Scope**

The principal areas of interest of the Section shall be, (1) to encourage the application of computer hardware, software and systems to statistical problems, (2) to encourage the application of statistical techniques to the design, maintenance and evaluation of computer hardware, software and systems, (3) to encourage the joint application of statistical techniques and computer technology to problems in other fields, and (4) to serve as the focal point for computer-oriented activities within the Association and for cooperation with computer-oriented organizations.

#### **Functions**

The Section will perform such functions as will support the areas of interest specified above. These will include: (1) sponsorship, or joint sponsorship with other organizations, of meetings, seminars or courses which involve statistics and computers; (2) planning, in cooperation with the General Program Committee, sessions on statistical computing and statistical computer science at annual or regional meetings of the Association; (3) sponsorship, or joint sponsorship with other organizations, of documentation of computer programs and algorithms of special interest to statisticians, of development of manuals and booklets and of publication of bibliographies, etc.

The Proceedings of the Section have been published since 1975. The Section established an ad hoc Committee on Evaluation of Statistical Program Packages, to initiate, carry out, and promote evaluation activities. In 1976 this committee was succeeded by the standing Committee on Statistical Software, with broader scope in promotion and guidance of related activities and less involvement in particular evaluations.

The interest in international cooperation, as shown at the 1967 meeting of the ISI and fostered by the 1969 Wisconsin conference, has grown through the continuing persistent efforts of many individuals but with especially vital participation by some from the start in 1967, leading to the establishment in 1977 of the International Association for Statistical Computing (IASC) as a new section of the ISI.

Highlighting of these activities is not intended to diminish the roles and importance of other efforts to promote and develop statistical computing, both in the U.S. and in other countries. Many individuals have been prominent and effective in more than one of these and other activities, contributing to increased communication and cooperation. This interaction extends to professional societies and meetings, especially through cooperative efforts in organizing sessions of mutual or complementary interest.

## **Influence on Development of Statistical Software**

Statistical computing may be viewed as consisting of the two parallel but overlapping areas of theory and application. Theory includes development of algorithms, evaluation and improvement of accuracy and efficiency, critical comparison of techniques, and statistical and mathematical computations in development of statistical theory. Products of theory are sometimes single-purpose, stand-alone computer programs not intended for general use or distribution but which eventually receive limited use

beyond the developer's environment. Other products include reports in the literature and presentations at scientific meetings. Application includes the assembling, packaging, documentation, distribution, and maintenance of a statistical program, group of programs, or system. The major product of application is the available (distributed) software. These areas are clearly related, and each profits from the work of the other. They also have nonoverlapping, distinct interests.

The recent history of statistical computing, in which it has received increased recognition and appreciation as a professional activity, has encouraged growth and production in theory and application, both separately and as they interact. Results from theory are more readily known and available to application, and needs of application receive acknowledged research from theory. The end result is the opportunity for continued, meaningful growth in the quality and availability of statistical software.

## Resources in Evaluating Statistical Software

Guidance for a prospective user in assessing the availability and quality of statistical software may be found from an increasing number of resources. Especially within the last five years, the improved climate for professional activities in statistical computing has led to progress in providing adequate answers to such questions as: What statistical programs (or systems) are available? Does the program do what I need, with verified accuracy? Can the program be used in my computer? Is the program adequately documented and easy to use? Is there help available when difficulties arise in the use of the program, especially if (when) errors are found?

A framework within which these and other questions can be effectively organized and addressed in the evaluation and improvement of statistical software was recently suggested by Francis (4), as he examined the interrelationships among the elements of the framework and summarized important results and developments in each area:

1. Evaluation of existing statistical software
  - Surveys and directories of software
  - Features of software (non-critical catalogues, critical reviews, experimental comparisons)
  - Publication of evaluations and recommendations, by individuals or societies
2. Development of professional standards
  - Methodology in development and evaluation of software

- Proposals for standards
- Role of statistical societies in standards
- Distribution of software by statistical societies

3. Developments in statistics
  - Current statistical practice reflected in software features
  - New developments in statistics to be included
4. Developments in computer science
  - Current abilities in numerical analysis, portability
  - New developments in general software, statistical languages, portability, data base management,
  - Numerical analysis
5. Goal: improvement of existing software, and development of new software.

Francis (4) provides 142 references organized by his more detailed display of this framework, and indicates plans to update and maintain this form of bibliography on evaluation of statistical software. An overlapping, more extensive bibliography of 261 references is also available (5).

Current scientific journals are not always appropriate or adequately prepared for publication of some, especially detailed, contributions to statistical computing. Nevertheless, important papers are found in *The American Statistician* (which from 1974 includes a Department on Statistical Computing) and the *Journal of the American Statistical Association*. The *Communications of the Association for Computing Machinery* (since 1975 through *Transactions on Mathematical Software*), publishes computing algorithms and related articles, as does *Applied Statistics*. *Communications in Statistics: Series B—Simulation and Computation*, *Technometrics*, and *Biometrics* occasionally contribute in this area. *Computer Programs in Biomedicine* is of potential interest to biostatisticians.

Two other resources, the most timely and important in the U.S. today, are the *Proceedings of the Statistical Computing Section of the American Statistical Association* (since 1975, but 1975 is out of print) and the *Proceedings of Computer Science and Statistics: Annual Symposia on the Interface* (since 1967; availability of various volumes is described in the introduction to the 11th Symposium, 1978).

A separate volume of the publication of the *International Statistical Institute*, 1979, contains information on capabilities and use of 46 major statistical computing packages, developed from a comprehensive questionnaire and poster session coordinated by Ivor Francis for the IASC program of the 41st Ses-

sion of the ISI, New Delhi (6). The summary matrix of program capabilities and features from that volume, together with an abbreviated description, is included here (Appendix I) with the permission of Professor Francis. It is planned that this information, together with other information to be collected by the IASC (7) will be provided in the future through an international information bank and information exchange on statistical computing software under Project TIESS (Technical Information Exchange on Statistical Software) of the IASC.

An earlier survey (8) of 56 publicly available statistical program packages provided a 119-page index containing a cross-reference listing of packages and general capabilities, a listing of developer's answers to selected questions concerning the design aspects of their packages, and abstracts of the package written by the developers (all available on microfiche). Matrices of general capabilities and selected questionnaire replies are included in Appendix II, with permission of the authors.

A third, recent comparison of nine major statistical packages by a featured matrix approach is based on user evaluation of developer's documentation and manuals, rather than direct response from developers to a questionnaire (9). This features matrix is included in Appendix III, with permission of the authors.

These summaries of program features are a useful resource to prospective users as a first step in evaluation of statistical program packages. Further developments in this area are anticipated and desirable. Evaluation criteria for statistical program packages are being developed and proposed (10, 11).

Critical review and comparison of programs, especially in terms of appropriate implementation of statistical methodology, accuracy, efficiency, and documentation, is perhaps more important and useful to a prospective (or current) user. The present, enlightened attitude of professional statisticians and societies has encouraged activity of this kind. Many recent papers of interest in this area are found in the Proceedings of the Section on Statistical Computing of the ASA and the Proceedings of Computer Science and Statistics: Annual Symposia on the Interface, and most are included in Francis' bibliographies (4, 5). Berk and Francis (12) and Muller (13) provide comprehensive, critical reviews of user manuals for two widely used statistical systems, BMDP (14) and SPSS (15).

Comparisons and critical reviews are needed, but they are produced in a dynamic environment and must be carefully examined for relevance to the currently available versions of programs, packages and documentation. Improvements are often concurrent

with evaluation. For example, the extensive reviews of SPSS and BMDP user guides have appeared just as the new BMDP-77 manual became available (16).

## Statistical Computing, Cancer Research, and Japan

Applications of biostatistics in cancer research will find useful features in many different statistical computing packages or systems. Few packages have been developed with special attention to the needs of statistical analysis of biological or medical data, a notable exception being the BMD-BMDP series of programs (14, 16, 17). The BMD programs were pioneers in the packaging, documenting, distribution, and maintenance of statistical systems, and their special emphasis and development of biomedical applications receives continued support from the National Institutes of Health. Inclusion in BMDP-77 of a program for log-linear analysis of contingency tables, and plans for early release of a program for stepwise multiple logistic analysis, reflect an ongoing effort to continue the special emphasis. Nevertheless, statistical computing in cancer research will undoubtedly find useful tools of different kinds in many software packages. Also, some of the needed tools will only be found in stand-alone, special-purpose programs, or not be found at all and have to be created ad hoc. For this latter purpose, the International Mathematical and Statistical Libraries (IMSL) subprogram library offers considerable resources (18).

Some of the major U.S. statistical computing systems are not available in Japan, and other system developers are interested in expanding their distribution to Japan. One system, Omnitab-78, claims to be multilingual to the extent that it could be provided in a Japanese version using the English alphabet (19). Portability of software may be an especially important question in transporting a U.S. system to a Japanese computer.

As a further guide to statistical computing in the U.S. with possible special relevance to Japan, the developers of the major U.S. systems included by Kohm, Ryan, and Velleman in their index of statistical software (8) were asked to briefly describe their present Japanese distribution (if any) and their interest in future distribution in Japan. With their reply they also sent current documentation for their systems, e.g., user guides or manuals, and primary journal articles which describe their systems. These materials, together with the Proceedings of the Section on Statistical Computing of the ASA for 1976 and 1977, the Proceedings of Computer Science and Statistics: Annual Symposia on the Interface for the

8th, 9th, 10th and 11th Symposia, 1975-1978, and Francis' bibliography (5) are provided for repository with a Japanese institution as an up-to-date information shelf on statistical computing in the U.S.

## Conclusion

Recent activities in statistical computing, both in the United States and through U.S. participants in international efforts, point to increasing progress, through statistical societies, meetings, and publications, toward research and products which benefit the users of computers in statistics. Users of both established statistical packages and of less widely publicized stand-alone programs can gain from the increased focusing of interest in more serious and visible work in standards, evaluations, critical comparisons, new methods, portability, distribution, and documentation.

Current directions suggested by recent meetings of societies and committees, and by their leaders and members, are ongoing work in evaluations and standards, with growing emphasis in interactive computing and in management of statistical data bases.

## Appendix I. Features Matrix by Francis.

This matrix displays responses by package developers to questions in four areas: capabilities, portability, ease of learning and using, and reliability. The complete forms of the abbreviated items shown in the matrix are found in the reference. Numeric responses are according to a 0-3 rating scheme: (0) no facilities in this area, or not intended as a goal; (1) a few functions in this area are present, or a minor goal or byproduct; (2) moderate capabilities, or a significant goal; (3) complete coverage of all aspects of this area, or one of the principal goals.

Alphabetic responses vary by the particular item, and are summarized here.

- 5: Program or package
  - A Single program
  - B Collection of (sub) programs
  - C Integrated but limited package
  - D Unified filing and analysis system
- 12: Generality of use
  - A Educational
  - B Special purpose program
  - C Package for specific tasks
  - D General system
- 13: Data set size
  - A Less than 1000 raw data records
  - B 1000-10,000 records
  - C 10,000-50,000 records
  - D Over 50,000 records

- 29: Extensibility of source code
  - A Private code
  - B Difficult modification
  - C Moderate modification
  - D Easy modification
- 48: Availability
  - A Private
  - B Export possible
  - C Use through service bureau, TS
  - D Export current
- 49: Installations
  - A One
  - B Two to five
  - C Six to nineteen
  - D Twenty or more
- 50: Computer makes
  - A One, machine dependent
  - B One, easily modified
  - C Two to four
  - D Five or more
- 51: Mini version
  - A Not possible
  - B Possible, with limitations
  - C Under development
  - D Exists
- 52: Core required
  - A 191 K
  - B 111 K to 190 K
  - C 49 K to 110 K
  - D 48 K or less
- 53: Batch/interactive
  - A Batch only
  - B Partially interactive
  - C Interactive only
  - D A and C
- 54: Statistics training (of user)
  - A Little
  - B One college-level course
  - C Two + courses
  - D M.A. +
- 55: Computer training (see 54)
- 56: Command language
  - A Numeric codes
  - B Alphabetic codes or short mnemonics
  - C English (or other) verbs and nouns
  - D English-like (or other) sentences
- 57: Documentation
  - A None
  - B General
  - C User manual and systems guide
  - D More than C
- 59: Maintenance
  - A Not done
  - B Limited
  - C Full; no further development
  - D Full plus development
- 60: Tested for accuracy
  - A Testing underway
  - B Developer tested; unpublished
  - C Developer tested; reproducible
  - D More than C

# PROGRAM CAPABILITIES

PROGRAM CAPABILITIES																		
A. GENERAL										B. PROCESSING AND . . . . .								
DESIGN GOALS										FILING								
PROGRAM OR PACKAGE	FILE MANIPULATION	EDITING	TABULATION	SAMPLE SURVEYS	STATIST. ANALYSIS	OTHER	GENERALITY OF USE	DATA SET SIZE	FLEXIBLE DATA INPUT	COMPLEX DATA STRUCTURES	HANDLE MISSING DATA	FILE MANIPULATION	FLEXIBLE RETRIEVAL	FILE MANIPULATION	FLEXIBLE OUTPUT			
	5	6	7	8	9	10	11	12		13	14	15	16	17	18	19		
<u>STATISTICAL ANALYSIS I</u>																		
BMDP 77	D	1	2	2	1	3	3	D	B	1	0	3	2	1	1			
CS	D	3	3	3	3	3	3	D	D	3	3	3	3	3	3			
DATATEXT	D	3	3	3	1	3	2	D	D	3	2	3	3	3	3			
GENSTAT	D	2	1	3	1	3	2	D	B	3	3	2	3	2	3			
HP STATPACKS	C	3	3	1	2	3	3	D	B	0	1	1	1	2	2			
MINITAB	C	1	2	2	0	2	2	D	B	2	1	1	1	1	1			
OMNITAB	D	3	2	3	1	3	3	D	A	3	2	2	1	3	1			
PSTAT	D	3	2	3	1	2	1	D	D	3	3	3	3	3	3			
OSIRIS III	C	2	2	3	0	2	1	C	D	2	2	3	2	2	2			
SAS 76.5	D	3	3	3	2	3	3	D	D	3	3	3	3	3	3			
SPSS	D	3	1	3	1	3	1	D	D	3	1	3	2	2	3			
SCSS	D	3	1	3	1	3	2	D	B	3	2	3	2	2	3			
SOUPAC	D	3	3	3	1	3	1	D	D	3	2	2	2	2	2			
TROLL	D	2	2	2	1	3	3	C	C	2	2	2	2	2	2			
<u>STATISTICAL ANALYSIS II</u>																		
AFSTAT	D	1	1	1	0	3	2	D	B	3	1	0	2	0	1			
ALLOC	B	0	0	2	0	3	0	C	A	2	1	0	0	0	2			
BIBLOS	D	2	2	0	0	3	2	D	D	2	1	0	0	0	0			
CADA	D	2	2	2	0	3	2	D	A	1	0	0	2	2	2			
GLIM	C	1	2	0	2	3	2	C	A	2	0	0	1	0	0			
LINWOOD	B	0	0	1	0	3	3	D	A	3	0	0	0	0	2			
MIT-SNAP	C	1	1	0	0	3	2	C	B	3	2	2	1	1	2			
ROSEPACK	B	2	2	0	3	3	3	D	D	0	3	0	3	0	3			
RUMMAGE	C	0	1	1	1	3	2	D	D	2	0	1	1	0	1			
STATSYS	B	1	0	0	0	2	0	C	B	1	1	1	0	1	1			
SURVO	C	1	1	3	0	3	1	D	D	2	2	2	2	2	3			
TSP	D	2	0	0	0	2	2	D	A	2	1	1	2	1	1			
TSP/DATATRA	D	1	2	0	0	3	3	C	D	3	2	1	3	2	3			
<u>DATA MANAGEMENT</u>																		
LEDA	B	3	3	3	3	1	0	C	D	3	2	3	3	3	3			
SIR	C	3	3	1	0	1	3	D	D	3	3	3	3	3	3			
WRAPS	A	2	2	2	0	1	3	C	D	3	3	2	3	3	3			
<u>EDITING</u>																		
CANEDIT	C	0	3	0	0	0	0	C	D	0	0	0	0	0	0			
CONCOR	D	3	3	0	0	0	0	C	D	3	3	3	3	3	3			
UNEDIT	C	0	3	0	0	0	0	C	D	2	2	2	0	0	0			
<u>TABULATION</u>																		
CENTS-AID	C	0	1	3	2	1	1	C	D	3	3	1	1	2	2			
COCENTS	C	1	1	3	0	1	0	C	D	3	3	3	0	0	3			
ISIS	C	3	2	2	0	0	1	C	C	3	2	2	2	2	0			
OSIRIS 2	C	1	1	3	0	1	0	C	D	2	2	1	1	1	3			
TAB 68	C	0	1	3	0	1	0	C	C	1	0	1	0	0	3			
TPL	C	0	1	3	3	2	2	C	D	3	3	3	3	3	3			
X-TALLY	C	0	1	3	1	1	0	C	D	2	0	2	2	0	1			
<u>SURVEY ANALYSIS</u>																		
CLUSTERS	B	0	0	0	2	1	0	C	D	0	0	1	0	0	1			
RGSP	B	2	2	3	3	1	1	C	D	3	2	3	2	2	3			
<u>SUBROUTINE LIBRARIES</u>																		
EISPACK	C	0	0	0	0	2	3	C	B	0	0	0	0	0	0			
IMSL	B	0	0	1	1	3	3	D	B	0	0	1	0	0	0			
NAG	B	0	0	1	0	2	2	D	A	3	0	3	3	0	3			

## ..... DISPLAYING DATA

209

PROGRAM CAPABILITIES (CONT.)

C. MATHEMATICAL ANALYSIS OF STATISTICAL DATA																
STATISTICAL ANALYSIS										SAMPLE SURVEY						
MULTIPLE REGRESSION	ANOVA/LINEAR MODEL	LINEAR MULTIVARIATE	MULTI-WAY TABLES	OTHER MULTIVARIATE	TIME SERIES	NONPARAMETRIC	EXPLORATORY	ROBUST	CURVE FITTING	BAYESIAN	ECONOMETRICS	COMPUTE ESTIMATES	COMPUTE VARIANCES	SELECT SAMPLES		
	30	31	32	33	34	35	36	37	38	39	40	41		42	43	44
<u>STATISTICAL ANALYSIS I</u>																
BMDP 77	3	2	3	3	2	0	2	1	1	3	1	0		2	0	1
CS	2	2	1	2	1	2	2	2	2	2	1	1		3	1	3
DATATEXT	3	3	1	2	2	0	2	0	0	0	0	0		0	0	0
GENSTAT	2	3	3	3	2	0	1	0	2	1	0	0		1	1	0
HP STATPACKS	3	3	1	2	2	2	3	1	0	3	0	1		1	0	0
MINITAB	2	1	0	0	0	2	1	0	1	0	0	0		0	0	1
OMNITAB	3	2	2	1	1	2	1	3	3	1	0	1		3	2	1
PSTAT	2	1	2	0	2	0	1	0	0	0	0	0		0	0	0
OSIRIS III	2	1	2	0	3	0	1	0	1	0	0	1		0	0	0
SAS 76.5	3	3	3	1	2	2	1	1	1	2	1	2		1	1	1
SPSS	1	1	1	1	2	0	3	0	0	0	0	0		1	0	1
SCSS	2	1	1	1	1	0	1	1	1	0	0	0		1	0	1
SOUPAC	2	1	2	1	2	2	1	0	0	1	0	2		1	0	1
TROLL	2	2	1	1	2	3	0	3	2	2	2	3		1	1	1
<u>STATISTICAL ANALYSIS II</u>																
AFSTAT	3	3	0	0	0	3	1	0	0	0	0	1		1	0	0
ALLOC	0	0	0	0	3	0	0	0	0	0	2	0		0	0	0
BIBLOS	3	0	3	0	3	0	0	0	0	0	0	0		0	0	0
CADA	2	1	0	0	0	0	0	2	1	0	3	0		2	0	0
GLIM	2	3	0	2	0	0	0	1	2	1	0	0		1	0	0
LINWOOD	3	0	0	0	0	0	0	2	0	3	0	0		0	0	0
MIT-SNAP	2	2	1	0	0	0	0	3	2	0	0	0		0	0	0
ROSEPACK	1	2	1	1	0	0	0	3	3	0	0	0		2	0	0
RUMMAGE	1	3	2	3	0	0	1	0	1	0	0	0		1	0	0
STATSYS	1	1	2	1	0	0	2	1	0	0	0	0		0	0	0
SURVO	2	1	2	1	2	3	1	0	0	0	0	1		0	0	0
TSP	2	0	1	0	0	1	0	0	0	2	0	3		0	0	0
TSP/DATATRAN	3	1	2	0	1	3	1	0	1	3	1	3		0	0	0
<u>DATA MANAGEMENT</u>																
LEDA	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
SIR	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
WRAPS	0	0	0	1	0	0	0	0	0	0	0	0		2	0	0
<u>EDITING</u>																
CANEDIT	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
CONCOR	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
UNEDIT	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
<u>TABULATION</u>																
CENTS-AID	0	0	0	0	0	0	0	0	0	0	0	0		2	0	0
COCENTS	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
ISIS	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
OSIRIS 2	0	0	0	0	0	0	0	0	0	0	0	0		2	0	0
TAB 68	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
TPL	0	0	0	0	0	0	0	0	1	0	0	0		1	1	0
X-TALLY	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
<u>SURVEY ANALYSIS</u>																
CLUSTERS	0	0	0	0	0	0	0	0	0	0	0	0		2	2	0
RGSP	0	0	0	0	0	0	0	0	0	0	0	0		2	2	0
<u>SUBROUTINE LIBRARIES</u>																
EISPACK	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
IMSL	3	3	3	3	3	3	3	1	2	3	1	3		2	2	1
NAG	2	1	2	1	0	0	1	0	0	3	1	1		1	0	3



	PORTABILITY				EASE OF LEARNING AND USING				RELIABILITY										
OTHER																			
SIMULATION	MATH. FUNCTIONS	OPERATIONS RESEARCH	AVAILABILITY	INSTALLATIONS	COMPUTER MAKES	MINI VERSION	CORE REQUIRED	BATCH/INTERACTIVE	STATISTICS TRAINING	NATURAL LANGUAGE	DOCUMENTATION	USER CONVENIENCES	MAINTENANCE	TESTED FOR ACCURACY					
	45	46	47		48	49	50	51	52	53		54	55	56	57	58		59	60
<u>STATISTICAL ANALYSIS I</u>																			
BMDP 77	2	1	0		D	D	D	D	B	A		B	B	D	D	2		D	C
CS	2	3	2		D	C	A	B	B	D		B	A	C	D	3		D	B
DATATEXT	0	0	0		D	D	C	B	A	A		B	A	C	C	3		D	D
GENSTAT	1	2	0		D	D	D	A	B	A		C	B	B	D	3		D	D
HP STATPACKS	1	3	2		D	D	A	D	D	C		B	A	B	D	3		D	A
MINITAB	1	1	0		D	D	D	D	D	D		A	A	C	D	3		D	C
OMNITAB	3	3	0		D	D	D	A	A	D		A	A	D	D	3		D	D
PSTAT	0	2	0		D	D	D	B	B	D		B	A	C	C	3		D	B
OSIRIS III	1	0	0		D	D	D	B	B	A		A	A	C	D	3		D	B
SAS 76.5	0	3	0		D	D	A	A	B	D		A	A	C	D	3		D	D
SPSS	1	0	0		D	D	D	D	B	A		B	A	B	D	3		D	B
SCSS	1	1	0		D	D	C	B	B	C		B	A	B	D	3		D	B
SOUPAC	1	2	1		D	D	C	B	B			B	B	B	C	3		C	B
TROLL	3	3	0		D	C	A	A	A	B		B	A	C	D	3		D	B
<u>STATISTICAL ANALYSIS II</u>																			
AFSTAT	1	0	0		D	B	C	B	A	D		B	A	B	B	1		D	A
ALLOC	0	0	0		D	C	C	B	B	A		B	A	A	D	2		D	C
BIBLOS	0	0	0		D	B	A	A	A	D		C	B	C	B	3		D	B
CADA	0	0	0		D	D	D	D	D	C		B	A	A	C	3		D	B
GLIM	2	0	0		D	D	D	B	C	D		C	B	B	C	3		D	C
LINWOOD	0	0	0		D	D	D	A	C	A		B	A	A	D	3		D	D
MIT-SNAP	1	1	0		D	C	A	A	B	D		A	A	C	B	2		C	B
ROSEPACK	1	3	0		D	D	D	D	A	D		A	A	B	D	3		D	D
RUMMAGE	0	0	0		D	D	D	C	B	B		D	A	C	B	2		D	D
STATSYS	0	0	0		D	C	C	B	B	A		C	A	A	B	1		A	B
SURVO	0	1	0		D	C	C	D	D	D		A	A	D	C	3		D	C
TSP	0	1	0		D	D	D	C	A	A		B	A	C	C	2		D	C
TSP/DATATRAN	2	3	0		C	C	C	A	A	D		B	A	C	C	3		D	A
<u>DATA MANAGEMENT</u>																			
LEDA	0	0	0		D	C	C	A	A	A		A	B	C	D	3		D	B
SIR	0	0	0		D	B	A	C	B	D		A	B	B	C	3		D	C
WRAPS	0	0	0		D	A	A	A	A	A		A	A	B	C	3		D	D
<u>EDITING</u>																			
CANEDIT	0	0	0		D	A	A	B	A	A		A	D	B	B	3		C	B
CONCOR	0	0	0		D	A	C	D	C	A		A	B	C	C	3		D	D
UNEDIT	0	0	0		D	A	B	D	D	B		A	A	B	B	3		D	B
<u>TABULATION</u>																			
CENTS-AID	0	0	0		D	D	C	B	B	A		A	A	D	C	3		D	C
COCENTS	0	0	0		D	D	D	D	D	A		A	A	B	C	2		D	D
ISIS	0	1	0		D	B	C	A	B	A		A	C	C	C	3		C	C
OSIRIS 2	0	0	0		B	B	B	A	A	B		B	B	C	C	3		D	C
TAB 68	0	0	0		D	C	A	A	B	A		A	A	C	B	3		D	D
TPL	0	0	0		D	D	C	A	A	A		A	A	C	D	3		D	D
X-TALLY	0	0	0		D	C	C	D	D	B		A	A	B	B	2		D	D
<u>SURVEY ANALYSIS</u>																			
CLUSTERS	0	0	0		B	B	B	B	C	A		C	B	A	B	2		C	B
RGSP	0	0	0		D	C	C	B	B	B		A	B	B	D	3		D	B
<u>SUBROUTINE LIBRARIES</u>																			
EISPACK	0	3	0		D	D	D	D	D	A		A	B	B	D	3		C	C
IMSL	3	3	2		D	D	D	D	D	D		C	C	A	C	2		D	B
NAG	3	3	2		D	D	D	D	D	D		B	B	B	D	3		D	B

## Appendix II. Features Matrices by Kohm, Ryan, and Velleman.

The capabilities and questions, which appear in abbreviated form in the "General Capabilities Listing" and the "Selected Questionnaire Replies" matrices, are found in their complete form on the microfiche.

The codes in the "General Capabilities Listing" were selected by the program developers in evaluation of the capabilities of the programs listed:

C Capability	The program or package has sufficient capabilities in this area to be considered as a feature.
L Limited	The program or package has some capabilities in this area, but they should be considered as limited.

D Documented	The feature can be easily accomplished using the documentation supplied with the program, but is not a standard ("built-in") option.
F Feasible	The feature can be accomplished by users with some expertise in the area.
I Insufficient	The program or package has insufficient capabilities in this area to be classified above.
or "Blank"	
P Planned	The program or package has either limited or no capability in this area, but program changes are planned to include this area in the "C" classification within one year.
O Obsolete	The program developer considers the program or this aspect of the program to be obsolete.

Programs may be listed in several categories. For example, a nonparametric analysis of variance program would be listed both in Analysis of Variance and in Nonparametric Statistics.

## GENERAL CAPABILITIES LISTING

9. Simple data description-----	1. File manipulation-----
10. Multiple linear regression-----	2. Variable labeling-----
11. Analysis of variance-----	3. Transformations-----
12. Analysis of covariance-----	4. Matrix manipulations-----
13. Canonical correlation & principal component-----	5. Sorting & matching-----
14. Factor analysis-----	6. Case selection/exclusion-----
15. Multivariate ANOVA & discriminant analysis-----	7. Consistency checking of data-----
16. Nonlinear regression-----	8. Auto error correction-----
17. Cluster analysis-----	27. Table printing (multi-way tables)-----
18. Simple analysis of multi-way tables-----	28. Graphics (versatile)-----
19. Loglinear analysis of multi-way tables-----	29. Reuse of computations in same run-----
20. Other analysis of categorical data-----	30. Storage of computations-----
21. Variance estimation (complex sample designs)-----	31. Run batch-----
22. Time series analysis-----	32. Run interactively-----
23. Data smoothing & curve fitting-----	33. Statistical level-----
24. Bayesian statistics-----	34. Computing level-----
25. Nonparametric statistics-----	
26. Random generation & Simulation-----	

Program	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	2	2	3	3	3	3	3	Abstract pages	I.D. #																
	6	5	4	3	2	1	0	9	8	7	6	5	4	3	2	1	0	9	1	2	3	4	5	6	7	8	7	8	9	0	1	2	3	4									
MIT-SNAP	l					c							l	c	c	l	l	c	l				l	c	c			b	a	b		18	-	20	1977-001								
AFSTAT	l	c		l	c			l	l		c		l	f	c			c	c	c	c		l	c	c	l	l	c	c	a	c	c		21	-	22	1977-002						
AQD				l	l			c	l	l	c	l	f	f	c			c	l	c	c		c	c	d	f	l	c	c	a	c	b		23	-	25	1977-003						
The IMSL Library	c	c	p	c	c	l	p	c	c		p	c	c	c	c	c		c	l	l	c		l	l	l	l	c	c		c	c		26	-	29	1977-004							
TABSTAT	l	c				l							c	c	f	c	c	l	c	c	l		c	c	c	c	f	n	a	c	c		30	-	31	1977-005							
TSP/DATATRAN	c	c	c	l	c				c	l	c	l	c	c	c	c		c	c	c	c		c	c	c	c	c	c	a	d	a		32	-	34	1977-006							
GLIM	l					c	c	f				c	c	c	c	l	c		l				l	c	c	f	c	c	b	c	b		35	-	35	1977-007							
RGSP						c	p	p	c														c	c	f	p	c	c		a	b		36	-	36	1977-008							
BMDP	c	c		f	p	c	c	c	c	c	c	c	c	c	c	c		c	l	c		l	c	c	c	c	c	c	a	b		37	-	41	1977-009								
BMD				f	c													c	l	l	c	l	f	c		l	l	l		a	b		42	-	44	1977-010							
WRAPS						l	l							l	c	c	l	l	c	c			c	c	c	c		c		c			45	-	46	1977-011							
LINWOOD and NONLINWOOD														l	l	c	c	c	f	f			l	c	c				b	b			47	-	48	1977-012							
SAS	f	l	f	f	l	c	f	f	c		c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	b	c	a		49	-	51	1977-013							
CS	l	c	c	c	c	l	c	l	c		c	f	c	l	f	c	c	c	c	c	c	c	c	c	c	c	c	c	n	a	c	b		52	-	55	1977-014						
P-STAT	p	l				c	p	c			c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	c	b	b	b		56	-	57	1977-015							
Minitab II	l	c			p	p	p						l	l	c	p	l	c	l	l	c		p	c	c	l	b	a	a			58	-	59	1977-016								
SPSS	l	c			p	p	p	c		p	c	l	l	c	c	c		l	c	c	f	l	c					a	a				60	-	61	1977-017							
RUMMAGE						f	l	c	f		l		c	c	l	f	l	c			f		l	l	c	c	c	c	b				62	-	63	1977-018							
HPSTAT PACS	f	c		c	f	c	p	l	p	p	f	c	c	c	c	c	l	c	f	c	l	f	c	c	c	c	n	a	a	b		64	-	65	1977-019								
CADA Monitor	p	p	c	p				p		f	l	c	c	c	c	c	p	p	c	c	c	c	c	c	c	c	p	n	a	b	a		66	-	67	1977-020							
IMPRESS***	o	c		d	d	c	d	c	d	c	f	f	d	d		c	c	c	c	d	d	c	l	c	d	c	d	n	a	a		68	-	68	1977-021								
SCSS						l	i	c								c	c	c	c	i		c	l	c		i	i	n	a			69	-	70	1977-022								
NTSYS	l							c	l	c				l	l	c	c	c					l	c	c			c	b			71	-	71	1977-023								
MATCAL														c	c	c	c	c					c	c	c			a	c			72	-	72	1977-024								
C-TAB II							c																c					c	b			73	-	73	1977-025								
EFAP, COFAMM, LISREL III							c			c	c												c		c	c		c	b			74	-	75	1977-026								
MULTIQUAL							c																c					c	b			76	-	76	1977-027								
MULTIVARIANCE										c	c	c	c	c	c	c		c					c		c	c		c	b			77	-	77	1977-028								
LOGO							c																c		c			c	b			78	-	78	1977-029								
NORMOG							c																					b	b			79	-	79	1977-030								
EXPAC														c														b	b			80	-	80	1977-031								
MOCA																												b	b			81	-	81	1977-032								
TSAR	c	f	f	f	f	f	f	c	f	f	c	f	c	d	c	c	c	c	f	l	c	c	l	c	l	c	l		c	a		82	-	82	1977-033								
GENSTAT	d	d	f	f	d	c	c	c	p	c	f	c	c	c	c	c	l	c	c	d	c	d	f	c	c	c	c	c	c	b		83	-	83	1977-034								
SOUPAC	l	l	f	l	f	c	l	l	c	c	c	l	c	c	c	c	l	c	c	d	c	d		l	l	c	c		c	b		84	-	85	1977-035								
ROSEPACK	l													c	c	c	d		c	c	c			l	c	c		a	c	c		86	-	86	1977-036								
STATII***	d	d	c	l		f	f	f	f	f	c	d	c	c	c	c	c	c	c	c	c			c	c	n	a	b	a		87	-	87	1977-037									
LOGLIN						c	p							l	c								c	c	c	a	c	b			88	-	88	1977-038									
Speakeasy	l	l		c	c	f	f	f	f	f	c	c	f	c	f	c	c	c	c	p	c	d	l	c	c	c	c	a	a		89	-	90	1977-039									
OMNITAB II	l	l				p	p	p	f	f	f	f	f	c	c	c	l	p	c	c	d	d	p	c	d	b	a	a			91	-	92	1977-040									
AESTH																						c	d	d	c	l	d	n	c	b		93	-	93	1977-041								
DATAPAC																												c	b			94	-	95	1977-042								
DATAPLOT	c	p		p	c		p	p	p	p	p	p	p	p	p	p	f	l	c	p	f	p	c		l	c	c		a	c		96	-	97	1977-043								
PROBPAC				f	f		f																c		c		n	a	c		98	-	99	1977-044									
ALSTAT	l	l				c	p						l	l	c	c	l	l	c		c	c	c	c	c	l		b	a		100	-	101	1977-045									
OSIRIS	l	c					p	c		c	d	c	c	l	c	c	c	c	c	l	c	d	c	l	p	c		c			102	-	105	1977-046									
CENTS-AID																		l	c	c			c		c			b	c			106	-	106	1977-047								
TSERIES							c																c	l	l		l	c		c	b		107	-	107	1977-048							
EXPLORII																								c	l	l			c	b		108	-	108	1977-049								
STATJOB										c		l	c	f	f	c	c	c	c	c	c	l	d	c	c	c	c	b	c	a		109	-	110	1977-050								
TPL																								c	c	c		c	a			111	-	111	1977-051								
XTABS																							c	c	c	c		c			112	-	113	1977-052									
GRAFSTAT																							c	c	c	c	f	f	c	d	f	l	c	f	n	a	b	a		114	-	114	1977-053
COCENTS	c	c		c	f	d				c	c	c		p	p	c	c	c	c	c	f	f	c	d	f	l	c	c	f	n	a	b	a		115	-	115	1977-054					
IDA																							l		c	c	d		c	b			116	-	117	1977-055							
Data-Text	c	c				l	c			c	c	f	c	i	c	c	c	c	c	c	c	c	c	l	c	l	c		a	a		118	-	119	1977-056								

Selected Questionnaire Replies

IDENTIFICATION	77- 1	77- 2	77- 3	77- 4	77- 5	77- 6	77- 7	77- 8	77- 9	77- 10	77- 11	77- 12	77- 13	77- 14	77- 15	77- 16	77- 17	77- 18	77- 19
2.1 RUN BATCH YES OR NO	n	y	y	n	y	y	n	n	n	n	n	n	n	y	n	n	n	n	y
2.2 INTERACTIVE A FULL/PROMPTS B FULL/NOPROMPTS C CONTROL CHECKING	b	a	a		a	a	b				c		b	a	b	b		c	a
2.3 AVAILABILITY A NO EXPORT B PASSIVE DISTRIB C ACTIVE DISTRIB D VIA TIMESHARING	b										b								
2.4 HOW AVAILABLE A RENTAL B ONE TIME COST C COMPUTER ASSOC	b	a	a	a		a	b	a	b	b		c	a	a		a	a		b
2.5 TRANSPORT METHOD A ONE BRAND COMPUT B PORTABLE LANGUAGE C DIFFERENT VERSIO	a		a			b	c	b	b	b	a	b	a	a		b	c	b	a
2.6 NUMBER OF SITES A 0 -10 B 11-99 C 100 OR MORE	a	a	a		a		a	b		c	a			a	b			b	a
2.7 DOCUMENTATION A PRIMER B REFERENCE C ON LINE HELP D IMPLEMENT GUIDE E OTHER	b	b	a	b	a	a	a	a	a	b	b	a	b	a	a	a	a	a	a
2.8 COMPILER & INTER A FORTRAN II B FORTRAN IV C NOT VERIFIED D PPORT E OTHER VERIFIER F EXTENDED FORT IV G COBOL H PL/1 I ASSEMBLER J BASIC K APL L ALGOL 60 M ALGOL 68 N OTHER O SPECIAL COMPILER	b	b		b		b	b	b	b	b		b		b	b	b		b	
2.10 DEVELOPMENT A CONTINUING B MAINTAINED ONLY C NO LONGER MAINT	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a	a	a	a	a
2.11 LOCAL MODIFICAT YES OR NO	n	n	n	y	y	y	y	y	y	y	n	y	y	y	n	y	y	y	y
2.12 LOCAL ADDITIONS YES OR NO	n	n	y	y	y	y	n	y	y	y	n	y	y	y	y	y	y	y	y
2.13 REVIEWED YES OR NO	n	n	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
3.1 RANK STRENGTHS A DATA ANALYSIS B MANIPULATION C SIMULATION D TEACHING E OTHER	1	1	1	1	2	2	1	2	1	1	3	1	1	1	2	2	1	1	1
3.2 BEST PROG DESC A MAIN PROGRAM B SEP MAIN PROGRAM C SPECIAL PURPOSE D INT. SYSTEM E SUBROUTINES	d	d	d	e	d	d		c	d	b		b		d	d	d	d	c	b
3.3 STATISTICAL EXP A NOVICE B NAIVE C MODERATE D ADVANCED	a								a	a		b				b	a	a	a
3.4 COMPUTER EXPR. A NOVICE B NAIVE C MODERATE D ADVANCED	b	c	b	c	c			c	b	b	c	b	a	b	b	a	a	b	b
3.5 GENERAL FIELDS A BIOLOGICAL SCIEN B SOCIAL SCIENCES C ENG & PHYSICAL D BUSINESS & ECON E GEN. STATISTICS F OTHER	e	e	e	e	e	d	e	e	e	e	e	e	e	e	e	e	e	e	e

## Selected Questionnaire Replies (continued)

IDENTIFICATION	77- 20	77- 21	77- 22	77- 23	77- 24	77- 25	77- 26	77- 27	77- 28	77- 29	77- 30	77- 31	77- 32	77- 33	77- 34	77- 35	77- 36	77- 37	77- 38
2.1 RUN BATCH YES OR NO	y	y	y	n	n	n	n	n	n	n	n	n	n	n	n	n	y	y	y
2.2 INTERACTIVE A FULL/PROMPTS B FULL/NOPROMPTS C CONTROL CHECKING	a	a	a														a	a	a
2.3 AVAILABILITY A NO EXPORT B PASSIVE DISTRIB C ACTIVE DISTRIB D VIA TIMESHARING	c	b	c	b	c	c	c	c	c	c	c	c	c	b	c	b	c	d	b
2.4 HOW AVAILABLE A RENTAL B ONE TIME COST C COMPUTER ASSOC	b	b	a	b	b	b	a	b	b	b	b	b	b	b	a	b	b		b
2.5 TRANSPORT METHOD A ONE BRAND COMPUT B PORTABLE LANGUAG C DIFFERENT VERSIO	c	a	a	a	c	c	c	c	c	c	c	c	b	a	c	a	b	c	b
2.6 NUMBER OF SITES A 0 -10 B 11-99 C 100 OR MORE	b	b	b	c	a	b	b	b	c	b	a	a	a	b	b	b	a	a	b
2.7 DOCUMENTATION A PRIMER B REFERENCE C ON LINE HELP D IMPLEMEN GUIDE E OTHER	b	a	a	b	b	b	b	b	a	b	b	b	b	a	a	b	c	a	a
2.8 COMPILER & INTER A FORTRAN II B FORTRAN IV C NOT VERIFIED D PPORT E OTHER VERIFIER F EXTENDED FORT IV G COBOL H PL/I I ASSEMBLER J BASIC K APL L ALGOL 60 M ALGOL 68 N OTHER O SPECIAL COMPILER	j	j	i			f	f	f	f	f	f			b	b	b	d	f	
2.10 DEVELOPMENT A CONTINUING B MAINTAINED ONLY C NO LONGER MAINT	a	a	a	a	a	b	a	a	a	a	a	a	b	b	a	a	a	a	
2.11 LOCAL MODIFICAT YES OR NO	y	y	y	y	y	y	n	y	y	y	y	y	y	y	n	y	y	n	y
2.12 LOCAL ADDITIONS YES OR NO	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y
2.13 REVIEWED YES OR NO	y	y	y	y	n	y	n	y	y	y	n	n	n	y	y	y	y	y	y
3.1 RANK STRENGTHS A DATA ANALYSIS B MANIPULATION C SIMULATION D TEACHING E OTHER	2	2	3	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
3.2 BEST PROG DESC A MAIN PROGRAM B SEP MAIN PROGRAM C SPECIAL PURPOSE D INT. SYSTEM E SUBROUTINES	d	d	d	d	e	a	b	a	a	a	a	a	a	a	a	d	c	a	a
3.3 STATISTICAL EXP A NOVICE B NAIVE C MODERATE D ADVANCED	b	a		c	a	c	c	c	c	c	b	b	b	c	c	c	c	b	c
3.4 COMPUTER EXPER. A NOVICE B NAIVE C MODERATE D ADVANCED	a	a		b	c	b	b	b	b	b	b	b	b	a	b	b	c	a	b
3.5 GENERAL FIELDS A BIOLOGICAL SCIEN B SOCIAL SCIENCES C ENG & PHYSICAL D BUSINESS & ECON E GEN. STATISTICS F OTHER	a	b		a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a

Selected Questionnaire Replies (continued)

IDENTIFICATION	77-39	77-40	77-41	77-42	77-43	77-44	77-45	77-46	77-47	77-48	77-49	77-50	77-51	77-52	77-53	77-54	77-55	77-56
2.1 RUN BATCH YES OR NO	y	n	n	n	y	n	n	n	n	n	n	n	n	n	y	n	y	n
2.2 INTERACTIVE																		
A FULL/PROMPTS	a				a										a		a	
B FULL/NOPROMPTS	b	b			b	b	b					b						
C CONTROL CHECKING	c		c									c						
2.3 AVAILABILITY																		
A NO EXPORT																		
B PASSIVE DISTRIB			b											b				
C ACTIVE DISTRIB	c	c		c	c	c	c	c	c	c	c	c	c		c	c	c	c
D VIA TIMESHARING	d							d	d			d			d			
2.4 HOW AVAILABLE																		
A RENTAL																		
B ONE TIME COST		b	b	b	b	b	b	b	b	b	b	b	b	b	a	b	b	a
C COMPUTER ASSOC															c			
2.5 TRANSPORT METHOD																		
A ONE BRAND COMPUT									a				a					
B PORTABLE LANGUAG		b	b	b	b	b	b		b	b	b			b	b	b	b	a
C DIFFERENT VERSIO	c							c							c	c	c	
2.6 NUMBER OF SITES																		
A 0 -10					a	a	a								a	a		
B 11-99			b	b					b	a	b	b					b	b
C 100 OR MORE	c	c						c					c			c		
2.7 DOCUMENTATION																		
A PRIMER	a	a		a	a	a		a	b	b	b	b	a	b		a		a
B REFERENCE	b	b		b	b		b	b	b	b	b	b		b	c	b		b
C ON LINE HELP	c				c	c											c	c
D IMPLEMEN GUIDE	d	d						d	d			d	e			d	d	e
E OTHER	e	e	e	e				e				e					e	
2.8 COMPILER & INTER																		
A FORTRAN II																		
B FORTRAN IV		b					b	b		b	b			b	b		b	b
ANSI																		
C NOT VERIFIED																		
D PFORT				d	d	d	d										c	
E OTHER VERIFIER																		
F EXTENDED FORT IV																		
G COBOL									g									
H PL/1							i	i				i						i
I ASSEMBLER																		
J BASIC			j															
K APL																		
L ALGOL 60																		
M ALGOL 68																		
N OTHER												n	n		n			
O SPECIAL COMPILER																		
2.10 DEVELOPMENT																		
A CONTINUING	a	a	a	a	a	a	a	a	a		b	b	b	a	a	a	a	a
B MAINTAINED ONLY																b		
C NO LONGER MAINT																		
2.11 LOCAL MODIFICAT																		
YES OR NO	n	y	y	y	y	y	y	y	n	y	y	y	y	y	n	y	y	y
2.12 LOCAL ADDITIONS																		
YES OR NO	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y
2.13 REVIEWED																		
YES OR NO	y	y	y	y	n	n	y	y	y	y	y	y	y	y	y	y	y	y
3.1 RANK STRENGTHS																		
A DATA ANALYSIS	1	1	1	1	1		1	1	1	1		1	2	2	1	3	1	1
B MANIPULATION	2	3		3	3			2	1			2	1	1	2	1	4	2
C SIMULATION		4		2											3	5		
D TEACHING	3	2	2	4	2					2	2	1	3		4	2	3	3
E OTHER				3	1											2		
3.2 BEST PROG DESC																		
A MAIN PROGRAM			a							a	a		a					
B SEP MAIN PROGRAM																		
C SPECIAL PURPOSE					c	c		d	d			d			c	c	b	d
D INT. SYSTEM		d															d	d
E SUBROUTINES				e														
3.3 STATISTICAL EXP																		
A NOVICE		a																a
B NAIVE		b	b				b		b						b			
C MODERATE	c	c		c	c	c		c		c	c	c	c	c		c	c	
D ADVANCED		d																
3.4 COMPUTER EXPR.																		
A NOVICE	a	a					a					a	a		a	b	b	a
B NAIVE			b							b	b							
C MODERATE				c	c	c		c	c					c				
D ADVANCED																		
3.5 GENERAL FIELDS																		
A BIOLOGICAL SCIEN	a		a						a	a	a	a	a	a	a		a	b
B SOCIAL SCIENCES	b		b					b	b	b	b	b	b	b		b	b	b
C ENG & PHYSICAL	c	c	c				c		c	c	c	c	c	c				
D BUSINESS & ECON	d		d				d		d	d	d	d	d	d		d	d	
E GEN. STATISTICS	e	e	e	e	e	e		e				e	e				e	
F OTHER		f																

## **Appendix III. Features Matrix by Hardy, Reynolds, and Kniefel.**

This features matrix, based on the authors' evaluation of developer documentation, emphasizes data management and transformation capabilities in addition to statistical features. The reference contains a glossary of terms used in the matrix, excluding statistical terms. Program packages included in this evaluation are: BMDP, DATA-TEXT, OMNITAB, OSIRIS II, P-STAT, SAS-76, SOUPAC, SPSS, and TSAR. Note that OMNITAB was reviewed, even though its successor, OMNITAB II, has been available for six years. In the matrix, X denotes the feature is present. Several minor corrections have been incorporated.

## COMPARISON OF STATISTICAL PACKAGE FEATURES

Input Data Capabilities	BMDP	DTXT	OMNI	OSIR	PSTA	SAS	SOUP	SPSS	TSAR
Types of Input Data									
Case by variable	X	X	X	X	X	X	X	X	X
Hierarchical records	X	X		X	X	X			
Variable length records						X			
Correlation matrices	X	X		X		X	X	X	X
Program Data Definition									
Mnemonic variable names	X	X			X			X	X
Column defined	X	X		X	X			X	X
FORTTRAN format	X		X				X	X	
Freefield						X		X	
Data Types									
Character < 5	X	X		X	X		X	X	X
Character > 4						X			
Multipunched (column binary)		X		X		X			
Multivalued observations		X		X					
Real binary	X	X	?		X	X	X	X	
Integer binary	X	X	?			X	X	X	
Packed decimal						X			
Zoned decimal						X			
Data Management									
Data Editing									
Input sequence			X					X	
Wild code check		X		X					
Range check	X	X		X					
Missing Values									
Automatic deletions	X	X		X	X	X	X	X	X
Pair-wise deletions	X	X		X	X	X	X	X	X
List-wise deletions	X	X		X	X	X	X	X	X
Checked in transformations	X	?	?	?	?	X		X	X
Data Transformation									
Recode statement	X	X		X			X	X	X
Character to numeric transform	X	X		X	X	X	X	X	X
Arithmetic computes	X	X	X	X	X	X	X	X	X
List functions	X	X	X	X	X	X	X	X	X
Crosscase transformations	X	X	X	X	X	X	X	X	X
Ranking	X	X	X	X	X	X	X	X	X
Standardization (Z scores)	X	X	X	X	X	X	X	X	X
Data aggregation	X	X		X	X		X	X	
Transpose data (e.g., case to variable)						X	X		
Contingent transformation	X	X	X	X	X	X	X	X	X
Case weighting	X	X	X	X	X	X	X	X	X
Sort functions			X	X	X	X	X	X	
Case Selection									
Random samples	X		X	X	X	X	X	X	X
Selective samples	X	X	X	X	X	X	X	X	X
Automatic storage of data subsets					X	X			
Package System File									
File Manipulation									
Save and process system files	X	X		X	X	X	X	X	X
Update system files	X	X		X	X	X	X	X	X
Add variables to file		X		X	X	X	X	X	X
Add cases to file		X		X	X	X	X	X	X
Merge files		X		X	X	X	X	X	X
File Interfaces									
Read other system files						X		X	
Write other system files		X							
Output									
Labeling									
Variables		X						X	
Values		X						X	
Other Output									
Data listing statement		X		X	X	X	X	X	X
Data to other (tape, etc.)		X	X	X	X	X	X	X	X
Matrix (correlation, etc.)		X		X			X	X	
Statistical Procedures									
Univariable Descriptive Measures									
Mean	X	X	X	X	X	X	X	X	X
Median	X	X	X	X	X	X	X	X	X
Mode	X	X	X	X	X	X	X	X	X
Variance	X	X	X	X	X	X	X	X	X
Standard deviation	X	X	X	X	X	X	X	X	X
Range	X	X	X	X	X	X	X	X	X
Frequency distribution	X	X	X	X	X	X	X	X	X
Histogram	X			X			X	X	
Contingency Table Analysis									
Row percent	X	X		X	X	X	X	X	X
Column percent	X	X		X	X	X	X	X	X
Cell percent of total	X	X		X	X	X	X	X	X
Expected values	X			X	X	X	X	X	X
Chi-square	X	X		X	X	X	X	X	X
Fisher's exact test	X	X		X	X	X	X	X	X
Yates' correction	X						X		
Non-Parametric Measures of Association									
Cramer's V	X	X		X			X	X	
Tau A	X	X		X					
Tau B	X	X		X	X			X	X
Tau C	X							X	
Gamma	X	X					X	X	
Somer's D									
Symmetric	X	X						X	
Asymmetric	X	X						X	
Lambda									
Symmetric	X	X					X	X	
Asymmetric	X	X					X	X	
Phi	X							X	

	BMDP	DTXT	OMNI	OSIR	PSTA	SAS	SOUP	SPSS	TSAR
Simple Correlation (Parametric)									
Scatter plots	X	X	X		X	X	X	X	X
Pearson's product moment	X	X	X	X	X	X	X	X	X
Eta2				X				X	
Simple regression coefficients	X	X	X	X	X	X	X	X	X
Sample Tests									
Parametric									
Student's T test	X	X		X		X	X	X	X
Related T test	X	X							
Non-Parametric									
Chi-square	X				X	X	X	X	X
Kolmogorov-Smirnov					X		X	X	X
Wilcoxon	X								
Runs test			X						
Tau A	X			X	X			X	
Concordance	X			X					
Mann-Whitney U	X						X		
Kruskal-Wallis Anova	X								
Friedman 2-way Anova	X								
Analysis of Variance									
Dimensions									
One-way	X	X		X	X	X	X	X	X
Two-way	X	X	X	X	X	X	X	X	X
Three-way	X	X	X	X	X	X	X	X	X
N-way	X	X	X	X	X	X	X	X	X
Estimation									
Unweighted means	X	X					X		X
Exact Anova	X		X	X	X	X			X
A priori contrasts									X
Posterior comparisons									X
Duncan multiple range test	X				X	X			X
Dunnett's T	X				X				X
Student-Newman-Keuls					X			X	X
Tukey							X		X
Tukey's alternative									X
Modified least significant difference									X
Scheffe							X		X
Repeated measures	X	X					X		X
Multiple Regression									
Variable entry									
Stepwise	X	X		X	X	X	X	X	X
Automatic polynomial	X		X				X		X
Parameter estimation technique									
Ordinary least squares	X	X	X	X	X	X	X	X	X
Weighted least squares									
Least squares estimates of nonlinear betas	X						X		
Assessment of resultant equation									
Residual printing	X	X	X	X	X		X	X	X
Residual plotting	X	X	X	X	X		X	X	X
Durbin-Watson	X		X	X	X	X	X	X	X
Multiple R	X	X	X	X	X	X	X	X	X
F-test for multiple R	X	X		X	X	X	X	X	X

	BMDP	DTXT	OMNI	OSIR	PSTA	SAS	SOUP	SPSS	TSAR
Parameters estimated									
Unstandardized beta	X	X	X	X	X	X	X	X	X
Standardized beta	X	X		X	X	X	X	X	X
Normalized beta					X				X
Regression through origin	X	X		X	X	X			X
T or F test for coefficients	X	X	X	X	X	X	X	X	X
Analysis of Covariance									
One-way	X	X		X	X	X		X	
N-way		X		X	X	X		X	
With multiple covariates	X			X	X	X		X	
Factor Analysis									
Factor structure estimation									
Principal components	X	X		X	X	X	X	X	X
Principal axis	X			X	X	X	X	X	X
More advanced techniques	X			X		X		X	X
Rotational methods									
Orthogonal									
Varimax	X	X		X	X	X	X	X	X
Other	X	X		X		X	X	X	X
Oblique				X				X	X
User supplied communalities	X					X	X	X	
Miscellaneous Multivariate Techniques									
Discriminant function	X			X	X	X	X	X	X
Canonical correlation	X			X	X	X	X	X	X
Probit	X				X	X	X		
Logit									
Cluster analysis	X			X		X	X		
AID (Automatic Interaction Detector)				X					
MCA (Multiple Classification Analysis)				X					
Spectral analysis	X					X	X		
Time series analysis	X					X	X		
Miscellaneous Mathematical Techniques									
Linear programming								X	
Matrix algebra operations			X				X	X	
Miscellaneous Scaling									
Nonmetric multidimensional scaling								X	
Guttman scaling					X	X		X	X
Roll call analysis				X					



## REFERENCES

1. Mann, N. R. Everything you always wanted to know about the history of computer science and statistics: Annual Symposia on the Interface — and more. In: *Proceedings of Computer Science and Statistics: Eleventh Annual Symposium on the Interface*. A. R. Gallant and T. M. Gerig, Eds., Institute of Statistics, North Carolina State University, Raleigh, N.C., 1978, p. 2.
2. Milton, R. C., and Nelder, J. A., Eds., *Statistical Computation*. Academic Press, New York, 1969.
3. *Statisticians and Others in Allied Professions*. The American Statistical Association, Washington, D.C., 1973.
4. Francis, I. The statistical profession and the quality of statistical software. *Bull. Int. Stat. Inst.*, in press.
5. Francis, I. A bibliography for statistical software: Reviews, evaluations, standards and improvements. Technical Report Series No. 1077/013, Department of Economic and Social Statistics, Cornell University, 1977.
6. Francis, I., Ed. A comparative review of statistical software. *International Association for Statistical Computing*, Voorburg, Netherlands, 1979.
7. Muller, M. E., and Wilkinson, G. N. A standardized form for describing programs, packages and systems for statistical applications. *Int. Stat. Rev.* 44: 349 (1976).
8. Kohm, R. F., Ryan, T. R., and Velleman, P. Index of available statistical software. *Proceedings of the Statistical Computing Section of the ASA*. American Statistical Association, Washington, D.C., 1977, p. 283.
9. Hardy, K. A., Reynolds, W. C., and Kniefel, D. R. Comparison of statistical packages: A features matrix approach. In: *Proceedings of Computer Science and Statistics: Tenth Annual Symposium on the Interface*. D. Hogben and D. W. Fife, Eds., NBS Special Publication 503, Washington, D.C., 1977, p. 342.
10. Francis, I., Heiberger, R. M., and Velleman, P. F. Criteria and considerations in the evaluation of statistical program packages. *Am. Statistician* 29: 52 (1975).
11. Velleman, P. F. and Welsch, R. E. Some evaluation criteria for interactive statistical program packages. *Proceedings of the Statistical Computing Section of the ASA*. American Statistical Association, Washington, D.C., 1975, p.10.
12. Berk, K. N., and Francis, I. A review of the manuals for BMDP and SPSS. *J. Am. Statist. Assoc.* 73: 65. (1978).
13. Muller, M. E. A review of the manuals for BMDP and SPSS. *J. Am. Statist. Assoc.* 73: 71 (1978).
14. Dixon, W. J. *BMDP Biomedical Computer Programs*. University of California Press, Berkeley, 1975.
15. Nie, N. H., et al. *SPSS: Statistical Package for the Social Sciences*. McGraw-Hill, New York, 1975.
16. Dixon, W. J., and Brown, M. B. *BMDP-77 Biomedical Computer Programs, P-Series*. University of California Press, Berkeley, 1977.
17. Dixon, W. J. *BMD Biomedical Computer Programs*. University of California Press, Berkeley, 1973.
18. *IMSL Library*. International Mathematical and Statistical Libraries, Houston, Texas, 1977.
19. Hogben, D. and Peavy, S. T. OMNITAB 78. In: *A Comparative Review of Statistical Software*. International Association for Statistical Computing. I. Francis, Ed., Voorburg, Netherlands, 1979.